

# Web Scraping Analysis for CSK 2025 Player Retention Recommendations

---

By: Rakesh roshan



# Objective

---

- CSK 2025 Player Retention Analysis
- Recommended Players for CSK Retention
- Reasons to retain

# Steps Performed in Analysis

---

- Data Scraping
- Data converting
- Data Cleaning
- Data Analysis
- Recommending players to retain
- Reasons for retaining

# Data Scraping

## Batsmen data

```
In [1]: import requests
        from bs4 import BeautifulSoup
        import pandas as pd

        url = "https://www.espncricinfo.com/records/trophy/averages-batting/indian-premier-league-117?team=4343"
        response = requests.get(url)

        soup = BeautifulSoup(response.content, 'html.parser')
        table = soup.find('table')

        data = []

        for row in table.find_all('tr'):
            row_data = []
            for col in row.find_all('td'):
                col_text = col.get_text()
                row_data.append(col_text)
            data.append(row_data)

        df = pd.DataFrame(data)
```

In [4]: df

Out[4]:

	0	1	2	3	4	5	6	7	8	9	10	11
0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	SR	100	50	0
1	P Amarnath	2008-2008	6	-	-	-	-	0.00	0.00	-	-	-
2	Akash Singh	2023-2023	6	-	-	-	-	0.00	0.00	-	-	-
3	MM Ali	2021-2024	48	40	4	853	93	23.69	136.26	-	3	4
4	S Anirudha	2008-2013	19	12	5	133	64	19.00	123.14	-	1	2
...	...	...	...	...	...	...	...	...	...	...	...	...

## Bowlers data

```
In [1]: import requests
        from bs4 import BeautifulSoup
        import pandas as pd

        url = "https://www.espncricinfo.com/records/trophy/averages-bowling/indian-premier-league-117?team=4343"
        response = requests.get(url)

        soup = BeautifulSoup(response.content, 'html.parser')
        table = soup.find('table')

        data = []

        for row in table.find_all('tr'):
            row_data = []
            for col in row.find_all('td'):
                col_text = col.get_text()
                row_data.append(col_text)
            data.append(row_data)

        df = pd.DataFrame(data)
```

In [2]: df

Out[2]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	Player	Span	Mat	Inns	Balls	Mdns	Runs	Wkts	BBI	Ave	Econ	SR	5	10	Ct	St
1	P Amarnath	2008-2008	6	6	132	-	236	7	2/29	33.71	10.72	18.8	-	-	2	-
2	Akash Singh	2023-2023	6	6	114	-	188	5	2/40	37.60	9.89	22.8	-	-	1	-
3	MM Ali	2021-2024	48	35	499	-	585	25	4/26	23.40	7.03	19.9	-	-	18	-
4	S Anirudha	2008-2013	19	-	0	-	-	-	-	0.00	-	-	-	-	11	-
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
99	S Vidyut	2008-2008	9	1	12	-	22	1	1/22	22.00	11.00	12.0	-	-	1	-

# Data Converting

```
import requests
from bs4 import BeautifulSoup
import pandas as pd

url = "https://www.espncricinfo.com/records/trophy/averages-batting/indian-premier-league-117?team=4343"
response = requests.get(url)

soup = BeautifulSoup(response.content, 'html.parser')
table = soup.find('table')

data = []

for row in table.find_all('tr'):
    row_data = []
    for col in row.find_all('td'):
        col_text = col.get_text()
        row_data.append(col_text)
    data.append(row_data)

df = pd.DataFrame(data)

file_path = 'D:\\CSK CRAP DATA\\Csk battingggg.xlsx'
df.to_excel(file_path, index=False)

print(f"DataFrame converted to Excel and saved to: {file_path}")
```

DataFrame converted to Excel and saved to: D:\CSK CRAP DATA\Csk battingggg.xlsx

```
In [1]: import requests
from bs4 import BeautifulSoup
import pandas as pd

url = "https://www.espncricinfo.com/records/trophy/averages-bowling/indian-premier-league-117?team=4343"
response = requests.get(url)

soup = BeautifulSoup(response.content, 'html.parser')
table = soup.find('table')

data = []

for row in table.find_all('tr'):
    row_data = []
    for col in row.find_all('td'):
        col_text = col.get_text()
        row_data.append(col_text)
    data.append(row_data)

df = pd.DataFrame(data)

file_path = 'D:\\CSK CRAP DATA\\Csk bowlers.xlsx'
df.to_excel(file_path, index=False)

print(f"DataFrame converted to Excel and saved to: {file_path}")
```

DataFrame converted to Excel and saved to: D:\CSK CRAP DATA\Csk bowlers.xlsx

# Data Cleaning

---

- Removed errors (like asterisks)
- Corrected players' names.
- Removed unnecessary text from values.
- Filled in missing data.
- Converted data into the correct formats.
- Removed retired players from the dataset.

# Data Cleaning

## Dataset before cleaning

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Player	Span	Mat	Inns	Balls	Mdns	Runs	Wkts	BBI	Ave	Econ	SR	5	10	Ct	St
P Amarnath	2008-2008	6	6	132	-	236	7	2/29	33.71	10.72	18.8	-	-	2	-
Akash Singh	2023-2023	6	6	114	-	188	5	2/40	37.60	9.89	22.8	-	-	1	-
MM Ali	2021-2024	48	35	499	-	585	25	4/26	23.40	7.03	19.9	-	-	18	-
S Anirudha	2008-2013	19	-	0	-	-	-	-	0.00	-	-	-	-	11	-
KB Arun Karthik	2010-2010	1	-	0	-	-	-	-	0.00	-	-	-	-	-	-
R Ashwin	2009-2015	97	94	2026	4	2180	90	3/16	24.22	6.45	22.5	-	-	20	-
KM Asif	2018-2021	3	3	49	-	93	4	2/43	23.25	11.38	12.2	-	-	-	-
S Badree	2014-2014	4	4	84	-	97	2	1/19	48.50	6.92	42.0	-	-	-	-
S Badrinath	2008-2013	95	-	0	-	-	-	-	0.00	-	-	-	-	16	-
GJ Bailey	2009-2010	4	-	0	-	-	-	-	0.00	-	-	-	-	-	-
L Balaji	2008-2010	29	29	562	-	809	31	5/24	26.09	8.63	18.1	1	-	5	-
SW Billings	2018-2019	11	-	0	-	-	-	-	0.00	-	-	-	-	6	-
DE Bollinger	2010-2012	27	27	576	3	693	37	4/13	18.72	7.21	15.5	-	-	8	-
DJ Bravo	2011-2022	116	113	2257	2	3147	140	4/42	22.47	8.36	16.1	-	-	59	-
DL Chahar	2018-2024	76	76	1597	6	2106	76	4/13	27.71	7.91	21.0	-	-	12	-
PP Chawla	2020-2020	7	7	126	-	191	6	2/33	31.83	9.09	21.0	-	-	-	-
DP Conway	2022-2023	23	-	0	-	-	-	-	0.00	-	-	-	-	8	-
SM Curran	2020-2021	23	22	450	-	672	22	3/19	30.54	8.96	20.4	-	-	8	-
TU Deshpande	2022-2024	31	31	671	-	1051	39	4/27	26.94	9.39	17.2	-	-	5	-
MS Dhoni	2008-2024	234	-	0	-	-	-	-	0.00	-	-	-	-	134	35
S Dube	2022-2024	41	3	18	-	50	1	1/14	50.00	16.66	18.0	-	-	14	-
F du Plessis	2012-2021	92	1	6	-	16	-	-	0.00	16.00	-	-	-	62	-
SP Fleming	2008-2008	10	-	0	-	-	-	-	0.00	-	-	-	-	2	-
A Flintoff	2009-2009	3	3	66	-	105	2	1/11	52.50	9.54	33.0	-	-	4	-
RD Gaikwad	2020-2024	66	-	0	-	-	-	-	0.00	-	-	-	-	36	-
C Ganapathy	2010-2010	1	1	6	-	13	-	-	0.00	13.00	-	-	-	-	-
RJ Gleeson	2024-2024	2	2	47	-	71	1	1/30	71.00	9.06	47.0	-	-	-	-
MS Gony	2008-2010	26	26	510	3	707	23	3/34	30.73	8.31	22.1	-	-	5	-
RS Hangargekar	2023-2023	2	2	36	-	60	3	3/36	20.00	10.00	12.0	-	-	-	-
Harbhajan Singh	2018-2019	24	23	455	1	582	23	3/20	25.30	7.67	19.7	-	-	8	-
JW Hastings	2014-2014	1	1	18	-	29	1	1/29	29.00	9.66	18.0	-	-	-	-
MS Gony	2008-2010	26	-	0	-	-	-	-	0.00	-	-	-	-	14	-

0	1	2	3	4	5	6	7	8	9	10	11
Player	Span	Mat	Inns	NO	Runs	HS	Ave	SR	100	50	0
P Amarnath	2008-2008	6	-	-	-	-	0.00	0.00	-	-	-
Akash Singh	2023-2023	6	-	-	-	-	0.00	0.00	-	-	-
MM Ali	2021-2024	48	40	4	853	93	23.69	136.26	-	3	4
S Anirudha	2008-2013	19	12	5	133	64	19.00	123.14	-	1	2
KB Arun Karthik	2010-2010	1	1	1	3	3*	0.00	60.00	-	-	-
R Ashwin	2009-2015	97	31	11	190	23	9.50	94.05	-	-	3
KM Asif	2018-2021	3	-	-	-	-	0.00	0.00	-	-	-
S Badree	2014-2014	4	-	-	-	-	0.00	0.00	-	-	-
S Badrinath	2008-2013	95	67	20	1441	71*	30.65	118.89	-	-	6
GJ Bailey	2009-2010	4	3	-	63	30	21.00	95.45	-	-	-
L Balaji	2008-2010	29	8	2	22	15	3.66	64.70	-	-	3
SW Billings	2018-2019	11	9	-	108	56	12.00	131.70	-	-	2
DE Bollinger	2010-2012	27	4	3	21	16*	21.00	91.30	-	-	-
DJ Bravo	2011-2022	116	75	35	1004	68	25.10	137.15	-	-	4
DL Chahar	2018-2024	76	12	6	66	39	11.00	124.52	-	-	3
PP Chawla	2020-2020	7	-	-	-	-	0.00	0.00	-	-	-
DP Conway	2022-2023	23	22	3	924	92*	48.63	141.28	-	-	2
SM Curran	2020-2021	23	15	4	242	52	22.00	142.35	-	-	1
TU Deshpande	2022-2024	31	2	2	-	0*	0.00	0.00	-	-	-
MS Dhoni	2008-2024	234	202	86	4669	84*	40.25	139.28	-	-	6
S Dube	2022-2024	41	39	7	1103	95*	34.46	159.16	-	-	3
F du Plessis	2012-2021	92	86	9	2721	96	35.33	131.44	-	-	5
SP Fleming	2008-2008	10	10	1	196	45	21.77	118.78	-	-	1
A Flintoff	2009-2009	3	3	1	62	24	31.00	116.98	-	-	-
RD Gaikwad	2020-2024	66	65	8	2380	108*	41.75	136.86	2	-	6
C Ganapathy	2010-2010	1	1	1	-	0*	0.00	0.00	-	-	-
RJ Gleeson	2024-2024	2	1	1	2	2*	0.00	100.00	-	-	-
MS Gony	2008-2010	26	9	5	50	15*	12.50	138.88	-	-	-

## Dataset after cleaning

Player	Mat	Inns	Balls	Mdns	Runs	Wkts	BBI	Ave	Econ	SR	5 Haul	10 Haul	Ct	St	Overs	No of years played
Moeen ali	48	35	499	0	585	25	4/26	23.4	7.03	19.9	0	0	18	0	83	4
Deepak Chahar	76	76	1597	6	2106	76	4/13	27.71	7.91	21	0	0	12	0	266	7
Devon Conway	23	0	0	0	0	0	0	0	0	0	0	0	0	8	0	3
Tushar Deshpande	31	31	671	0	1051	39	4/27	26.94	9.39	17.2	0	0	5	0	112	5
MS Dhoni	234	0	0	0	0	0	0	0	0	0	0	0	134	35	0	15
Shivan Dube	41	3	18	0	50	1	1/14	50	16.66	18	0	0	14	0	3	3
Ruturaj Gaikwad	66	0	0	0	0	0	0	0	0	0	0	0	36	0	0	5
Richard Gleeson	2	2	47	0	71	1	1/30	71	9.06	47	0	0	0	0	8	3
Rajvardhan Hangargekar	2	2	36	0	60	3	3/36	20	10	12	0	0	0	0	6	3
Ravindra Jadeja	172	159	2955	2	3727	133	5/16	28.02	7.56	22.2	1	0	85	0	493	13
Daryl Mitchell	13	2	36	0	70	1	1/18	70	11.66	36	0	0	9	0	6	1
Mukesh Choudhary	14	14	279	0	451	16	4/46	28.18	9.69	17.4	0	0	3	0	47	4
Mustafizur Rahman	9	9	206	1	318	14	4/29	22.71	9.26	14.7	0	0	1	0	34	2
Matheesha Pathirana	20	20	451	0	592	34	4/28	17.41	7.87	13.2	0	0	6	0	75	3
Dwaine Pretorius	7	7	150	0	238	6	2/30	39.66	9.52	25	0	0	3	0	25	3
Ajinkya Rahane	27	0	0	0	0	0	0	0	0	0	0	0	9	0	0	3
Rachin Ravindra	10	2	12	0	7	0	0	0	3.5	0	0	0	8	0	2	3
Sameer Rizvi	8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2
Mitchell Santner	18	18	366	0	422	15	2/13	28.13	6.91	24.4	0	0	6	0	61	7
Simarjeet Singh	10	10	180	0	259	9	3/26	28.77	8.63	20	0	0	3	0	30	3
Shardul Thakur	57	56	1203	2	1823	60	3/28	30.38	9.09	20	0	0	17	0	201	4
Maheesh Theekshana	27	27	624	1	797	25	4/33	31.88	7.66	24.9	0	0	4	0	104	3

Player	From Date	To Date	Total Seaso	Mat	Inns	NOT OUT	Runs	Average	SR	100	50	Duck out	Total years	Highest Scd
Ajinkya Rahane	2023	2024	2	27	23	1	568	25.81	147.53	0	2	0	1	71
Daryl Mitchell	2024	2024	1	13	13	2	318	28.9	142.6	0	2	0	0	63
Deepak Chahar	2018	2024	7	76	12	6	66	11	124.52	0	0	3	6	39
Devon Conway	2022	2023	2	23	22	3	924	48.63	141.28	0	9	2	1	92
Matheesha Pathirana	2022	2024	3	20	0	0	0	0	0	0	0	0	2	0
Mahesh Theekshana	2022	2024	3	27	2	1	7	7	100	0	0	1	2	7
Mitchell Santner	2019	2024	6	18	11	5	70	11.66	98.59	0	0	1	5	22
Moeen Ali	2021	2024	4	48	40	4	853	23.69	136.26	0	3	4	3	93
MS Dhoni	2008	2024	17	234	202	86	4669	40.25	139.28	0	22	6	16	84
Mukesh Choudhary	2022	2024	3	14	2	1	6	6	100	0	0	0	2	4
Mustafizur Rahman	2024	2024	1	9	0	0	0	0	0	0	0	0	0	0
Prashant Solanki	2022	2022	1	2	0	0	0	0	0	0	0	0	0	0
Rachin Ravindra	2024	2024	1	10	10	0	222	22.2	160.86	0	1	1	0	61
Ravindra Jadeja	2012	2024	13	172	125	57	1897	27.89	136.18	0	3	5	12	62
Ruturaj Gaikwad	2020	2024	5	66	65	8	2380	41.75	136.86	2	18	6	4	108
Richard Gleeson	2024	2024	1	2	1	1	2	0	100	0	0	0	0	2
Rajvardhan Hangarg	2023	2023	1	2	0	0	0	0	0	0	0	0	0	0
Shivam Dube	2022	2024	3	41	39	7	1103	34.46	159.16	0	8	3	2	95
Sudeep Tyagi	2009	2010	2	14	3	2	3	3	75	0	0	1	1	3
Sameer Rizvi	2024	2024	1	8	5	1	51	12.75	118.6	0	0	1	0	21
Simarjeet Singh	2022	2024	3	10	3	2	7	7	87.5	0	0	0	2	3
Shardul Thakur	2018	2024	7	57	13	7	67	11.16	121.81	0	0	2	6	17
Tushar Deshpande	2022	2024	3	31	2	2	0	0	0	0	0	0	2	0

# Data Analysis

Used a pivot table to analyze the top performers in both batting and bowling

Row Labels	Sum of Mat	Sum of Inns	Sum of Runs	Sum of Wkts	Sum of Econ	Sum of Ave	Sum of No of years played
AM Rahane	27	0	0	0	0	0	3
D Pretorius	7	7	238	6	9.52	39.66	3
DJ Mitchell	13	2	70	1	11.66	70	1
DL Chahar	76	76	2106	76	7.91	27.71	7
DP Conway	23	0	0	0	0	0	3
M Pathirana	20	20	592	34	7.87	17.41	3
M Theekshana	27	27	797	25	7.66	31.88	3
MJ Santner	18	18	422	15	6.91	28.13	7
MM Ali	48	35	585	25	7.03	23.4	4
MS Dhoni	234	0	0	0	0	0	15
Mukesh Choudhary	14	14	451	16	9.69	28.18	4
Mustafizur Rahman	9	9	318	14	9.26	22.71	2
R Ravindra	10	2	7	0	3.5	0	3
RA Jadeja	172	159	3727	133	7.56	28.02	13
RD Gaikwad	66	0	0	0	0	0	5
RJ Gleeson	2	2	71	1	9.06	71	3
RS Hangargekar	2	2	60	3	10	20	3
S Dube	41	3	50	1	16.66	50	3
Sameer Rizvi	8	0	0	0	0	0	2
Simarjeet Singh	10	10	259	9	8.63	28.77	3
SN Thakur	57	56	1823	60	9.09	30.38	4
TU Deshpande	31	31	1051	39	9.39	26.94	5
Grand Total	915	473	12627	458	151.4	544.19	99

PivotTable Fields

Choose fields to add to report:

Search

- ☒ Ave
- ☒ Econ
- ☐ SR
- ☐ 5 Haul
- ☐ 10 Haul
- ☐ Ct
- ☐ St
- ☐ Overs
- ☒ No of years played

More Tables...

Drag fields between areas below:

Filters	Columns
	<div>Σ Values</div>
Rows	Σ Values
Player	Sum of Mat
	Sum of Inns
	Sum of Runs
	Sum of Wkts

Row Labels	Sum of Inns	Sum of Runs	Sum of Average	Sum of SR	Sum of Highest Score	Sum of Total years played
Ajinkya Rahane	23	568	25.81	147.53	71	1
Daryl Mitchell	13	318	28.9	142.6	63	0
Devon Conway	22	924	48.63	141.28	92	1
DL Chahar	12	66	11	124.52	39	6
M Pathirana	0	0	0	0	0	2
M Theekshana	2	7	7	100	7	2
Mitchell Santner	11	70	11.66	98.59	22	5
MM Ali	40	853	23.69	136.26	93	3
MS Dhoni	202	4669	40.25	139.28	84	16
Mukesh Choudhary	2	6	6	100	4	2
Mustafizur Rahman	0	0	0	0	0	0
Prashant Solanki	0	0	0	0	0	0
R Ravindra	10	222	22.2	160.86	61	0
Ravindra Jadeja	125	1897	27.89	136.18	62	12
RD Gaikwad	65	2380	41.75	136.86	108	4
RJ Gleeson	1	2	0	100	2	0
RS Hangargekar	0	0	0	0	0	0
S Dube	39	1103	34.46	159.16	95	2
S Tyagi	3	3	3	75	3	1
Sameer Rizvi	5	51	12.75	118.6	21	0
Simarjeet Singh	3	7	7	87.5	3	2
SN Thakur	13	67	11.16	121.81	17	6
TU Deshpande	2	0	0	0	0	2
Grand Total	593	13213	363.15	2226.03	847	67

PivotTable Fields

Choose fields to add to report:

Search

- ☐ NOT OUT
- ☒ Runs
- ☒ Average
- ☒ SR
- ☐ 100
- ☐ 50
- ☐ Duck out
- ☒ Total years played
- ☒ Highest Score

More Tables...

Drag fields between areas below:

Filters	Columns
	<div>Σ Values</div>
Rows	Σ Values
Player	Sum of Inns
	Sum of Runs
	Sum of Average
	Sum of SR



# Recommended players to retain

## Players Retention Rules

There should be a total of 5 players

Out of those, 3 must be Indian.

The remaining 2 should be foreign players.

## Recommended Players to retain

	Batsmen/All Rounder	Bowler
Overseas	Devon conway	M Pathirana
Indians	Ruturaj Gaikwad and Ravindra Jadeja	Deepak Chahar

### Bowlers' contribution to the team

Player Name	Innings	Total Overs	Wickets	Economy	Average	Strike Rate
Ravindra jadeja	159	493	133	7.56	28.02	22.2
Deepak chahar	76	266	76	7.91	27.71	21
Shardul thakur	56	201	60	9.09	30.38	20
Tushar deshpane	31	112	39	9.39	26.94	17.2
Matheesha pathirana	20	75	34	7.87	17.41	13.2
Moeen Ali	35	83	25	7.03	23.4	19.9
Maheesh Thekshana	27	104	25	7.66	31.88	24.9
Mukesh Choudhary	14	47	16	9.69	28.18	17.4

### Batsmen contribution to the team

Player Name	Innings	Total Runs	Average	Strike rate
MS Dhoni	202	4669	40.25	139.28
Ruturaj gaikwad	65	2380	41.75	136.86
Ravindra Jadeja	125	1897	27.89	136.18
Shivam dube	39	1103	34.46	159.16
Devon Conway	22	924	48.63	141.28
Moeen ali	40	853	23.69	136.26
Ajinkya Rahane	23	568	25.81	147.53
Daryl Mitchell	13	318	28.9	142.6
Rachin ravindra	10	222	22.2	160.86

# Reasons for retaining

Ruturaj Gaikwad (Indian - Batsman, Future Captain)

**Performance:** 924 runs in 22 innings

**Average:** 48.63

**Strike Rate:** 141.28

**Special Note:** Conway is a dependable left-handed batsman.

**Reason:** He is a consistent performer with a strong average and strike rate, making him reliable at the top of the order.

Devon Conway (Overseas - Batsman)

**Performance:** 924 runs in 22 innings

**Average:** 48.63

**Strike Rate:** 141.28

**Special Note:** Conway is a dependable left-handed batsman.

**Reason:** He is a consistent performer with a strong average and strike rate, making him reliable at the top of the order.

Matheesha Pathirana (Overseas - Bowler)

**Performance:** 34 wickets in 20 matches

**Economy:** 7.87

**Strike Rate:** 13.2

**Special Note:** Pathirana is especially effective at bowling in the death overs.

**Reason:** His ability to take wickets and control runs during crucial moments makes him a key bowler for the team.

Deepak Chahar (Indian - Bowler)

**Performance:** 76 wickets in 76 matches

**Economy:** 7.91

**Strike Rate:** 21

**Reason:** Chahar's strength lies in taking early wickets, especially in the powerplay, which gives the team a strong start with the ball.

Ravindra Jadeja (Indian - All-rounder)

Performance:

**Batting:** 1897 runs in 125 innings

**Bowling:** 133 wickets in 159 innings

**Reason:** Jadeja's ability to perform well with both bat and ball makes him an essential all-rounder for the team

*Thank  
you!*